

Entropy-based inhomogeneity detection in porous media

Patricia Alonso Ruiz and Evgeny Spodarev

Abstract

We study a change-point problem for random fields based on a univariate detection of outliers via the 3σ -rule in order to recognize inhomogeneities in porous media. In particular, we focus on fibre reinforced polymers modeled by stochastic fibre processes with high fibre intensity and search for abrupt changes in the direction of the fibres. As a measure of change, the entropy of the directional distribution is locally estimated within a window that scans the region to be analyzed.

Keywords: inhomogeneity detection, entropy, fibre process, change-point problem, Boolean model.

1 Introduction

Lightweight materials are highly demanded in many industrial applications, for instance in automobile, aerospace or wind turbine construction. Fibre reinforced polymers (FRP) constitute an important class of such materials, whose macroscopic properties are directly influenced by their microstructure, and in particular by the orientation of the fibres. With the aim of gaining a better understanding of this relation, methods of non-destructive characterization such as the analysis of 3-dimensional images by micro-computed tomography (μ CT), combined with stochastic microstructure modeling, are currently being investigated [5, 7, 8].

In the compression moulding process of a FRP, the fibres order themselves inside the raw material as a result of mechanical pressure. During this process, deviations from the requested direction may occur, creating undesirable fibre clusters and/or deformations. This kind of inhomogeneities are characterized by abrupt changes in the direction of the fibres and its detection is subject of study in the change-point analysis. In particular in this paper, we are interested in developing a method that is sensitive to changes and at the same time is considerably fast. In order to reduce the complexity of the problem carried out by spatial data, we investigate the random field generated by the entropy of the directional distribution of the fibres estimated in a moving scanning window that runs over the observed piece of material. In this way, the initial spatial change-point problem becomes a univariate detection of outliers on a random field.

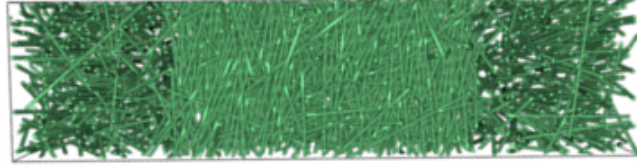


Figure 1: A simulated fibre structure (courtesy of Katja Schladitz, ITWM Kaiserslautern).

The entropy is estimated in a non-parametric manner by means of the entropy estimator introduced in [1]. Due to its asymptotic normality, see (2.6), the 3σ -rule yields a criterion to identify outliers. Searching for points above or below the given threshold depends on the specific application and both procedures are completely analogous. Moreover, this procedure may be applicable not only to the direction but to any other characteristic of interest, for instance the fibre length or the fibre curvature.

It is important to remark at this point that the use of the 3σ -rule is justified by the asymptotic normality of the entropy estimator. As a result, the proposed method is reliable on structures with high fibre intensity. In the simulation study presented in Section 4 we have therefore generated structures with an intensity of 5 fibres per unit volume. If we consider for instance a cube of side length $10\ \mu\text{m}$ filled with fibres of radius around $1.5\ \mu\text{m}$ and length $10\ \mu\text{m}$, this corresponds to a fibre volume ratio of 35%, a usual percentage in high-densely FRPs.

We have worked deliberately with simulated data for several reasons: on the one hand, the inhomogeneity region is known a priori and it can thus easily be compared with the output. This ensures an accurately testing of the performance of the detection algorithm. On the other hand, the segmentation of 3-dimensional images of high-intensity fibre structures is still difficult and expensive.

In connection with this kind of detection problems one can find in the literature parametric scan-statistics [12], multivariate cumulative sums (cusum) methods [3] and change-point analysis [4] for spatial data.

The fibre structures modeled in the present paper correspond to those having relatively short and rigid fibres that will be assumed to have the same length. More precisely, we consider the *Boolean model*

$$\Phi_\lambda = \bigcup_{Y_i \in \Pi_\lambda} (F_{Y_i} + Y_i),$$

where $\Pi_\lambda = \{Y_i\}_{i \geq 1} \subseteq \mathbb{R}^3$ is a *homogeneous Poisson point process* of intensity $\lambda > 0$ and F_{Y_i} denotes an independent copy of a line segment F with a given length $\ell > 0$. Each point Y_i corresponds to the center of a fibre F_{Y_i} and thus Φ_λ is a random set that models a system of independent fibres. To each center Y_i a mark ξ_i is attached that represents the (random) unit direction vector of the fibre F_{Y_i} . These marks are assumed to be i.i.d. random variables independent of the location of the points Y_i , and

their distribution is supposed to have a continuous density f with respect to a spherical surface measure.

Based on the marked Poisson point process (MPPP) $\Psi_\lambda := \{(Y_i, \xi_i)\}_{i \geq 1}$, a non-parametric estimator of the entropy of the directional distribution has been introduced in [1]. Although the Poisson assumption is too strong to actually be considered in applications, we have focused on reproducing the property of high fibre volume ratio. Applying our methods to more realistic models such as hard-core processes is still a matter of future research.

The paper is organized as follows: in Section 2, definitions and theoretic results concerning the estimation of entropy for the MPPP Ψ_λ are presented. Section 3 describes the proposed method of detection of inhomogeneities in the direction of the fibres and investigates the optimal size relationship between the inhomogeneity region and the scanning window size. Section 4 is devoted to a simulation study of the estimators and the detection algorithm. The paper finishes in Section 5 with a discussion of possible limitations of the method which are object of further research.

2 Non-parametric entropy estimation.

The key element in the proposed method to detect inhomogeneities in a FPR is the use of the *entropy* of the directional distribution of the fibres composing it. The concept of entropy was introduced by Shannon in [11] to provide a way of measuring the closeness of a distribution to the uniform distribution. Given a random variable ξ that takes values on an abstract measurable space (M, ν) and has distribution density $f_\xi: M \rightarrow \mathbb{R}$ with respect to the measure ν , its entropy is defined as

$$\mathcal{E}_{f_\xi} := -\mathbb{E}[\log f_\xi(\xi)] = -\int_M \log f_\xi(x) f_\xi(x) \nu(dx).$$

As mentioned in the introduction, the fibre structures being considered here are modeled by the MPPP

$$\Psi_\lambda = \{(Y_i, \xi_i)\}_{i \geq 1} \subseteq \mathbb{R}^3 \times S^2, \quad \lambda > 0, \quad (2.1)$$

where S^2 denotes the unit sphere in \mathbb{R}^3 . The Euclidean space will be equipped with the Lebesgue measure $\text{vol}(dy) := dy$, and S^2 with the surface-area measure $\omega(\cdot)$. The corresponding σ -algebras of Borel sets will be denoted by $\mathcal{B}(\mathbb{R}^3)$, respectively $\mathcal{B}(S^2)$.

The present section summons up the theoretic results concerning density and entropy estimation of the directional distribution f in this particular 3-dimensional setting. We refer to [1] for the more general case of MPPPs with manifold-valued marks.

2.1 Kernel estimator of the directional distribution density

Let $B \in \mathcal{B}(\mathbb{R}^3)$ denote the observation window (image, piece of material) that provides the data. For simplicity, one can think of $B = [0, b]^3$ for some $b > 0$. The estimator of

the directional density $f: S^2 \rightarrow \mathbb{R}$ on B proposed in [1] is given by

$$\hat{f}_B(\eta) := \frac{1}{\lambda \text{vol}(B)} \sum_{i \geq 1} \frac{\mathbb{1}_{\{Y_i \in B\}}}{h^2 \theta_\eta(\xi_i)} K\left(\frac{d_g(\eta, \xi_i)}{h}\right), \quad \eta \in S^2, \quad (2.2)$$

where $h > 0$ denotes the *bandwidth*, $\theta_\eta(\xi_i) = |\sin \arccos \langle \eta, \xi_i \rangle| / \arccos \langle \eta, \xi_i \rangle$ is the so-called *volume density function*, $K: \mathbb{R}_+ \rightarrow \mathbb{R}$ is a *kernel function*, and $d_g(\eta, \xi_i) = \arccos \langle \eta, \xi_i \rangle$ is the *geodesic distance* between $\eta, \xi_i \in S^2$. Here, $\langle \cdot, \cdot \rangle$ denotes the standard scalar product on \mathbb{R}^3 .

Under the proper choice of a kernel K and a sequence of bandwidths $\{h_n\}_{n \geq 1}$ (see [1, Corollary 3.2] for details), this estimator is L^2 -consistent, i.e.

$$\mathbb{E}[\|\hat{f}_{B_n} - f\|_2^2] \xrightarrow{n \rightarrow \infty} 0,$$

where $\{B_n\}_{n \geq 1} \subseteq \mathbb{R}^3$ is a sequence of regularly growing Borel sets, for instance $B_n = [0, b_n]^3$, with $b_n \rightarrow \infty$, and $\|f\|_2^2 = \int_{S^2} f^2(\xi) \omega(d\xi)$. In addition, if the observation windows B_n are large enough, \hat{f}_{B_n} is almost surely consistent as well (c.f. [1, Theorem 3.8]). Moreover, it is possible to give an asymptotically optimal bandwidth

$$h_{opt} = \left(\frac{C_\theta K_0^2}{2C_2^2 K_2^2 \lambda \text{vol}(B)} \right)^{1/6},$$

where the constants $C_\theta, K_0, K_2, C_2 > 0$ depend on the volume density function θ_η , the kernel K and the density f .

2.2 Entropy estimator

The non-parametric estimator of the directional distribution f in the observation window B is defined as

$$\hat{\mathcal{E}}_f(B) := -\frac{1}{\lambda \text{vol}(B)} \sum_{i \geq 1} \mathbb{1}_{\{Y_i \in B\}} \log \hat{f}_{B'+Y_i}(\xi_i), \quad (2.3)$$

where $B' \subseteq B$ is the sub-window in which the density f is estimated and $B' + y := \{x + y, x \in B'\}$ denotes the translation of B' by $y \in \mathbb{R}^3$. The additional window B' is introduced for the purpose of notation and it is only relevant when studying the asymptotic normality of the estimator. For the purpose of estimation and consistency there are no restrictions on it and one can assume $B' = B$.

Under the proper assumptions on the kernel K , the sequence of bandwidths $\{h_n\}_{n \geq 1}$ and some integrability properties of f (see [1, Theorem 4.1] for details), this estimator is L^2 -consistent, i.e.

$$\mathbb{E}[|\hat{\mathcal{E}}_f(B_n) - \mathcal{E}_f|^2] \xrightarrow{n \rightarrow \infty} 0,$$

for a sequence of regularly growing Borel sets $\{B_n\}_{n \geq 1} \subseteq \mathbb{R}^d$. The bias of $\hat{\mathcal{E}}_f(B_n)$ is controlled by the bandwidth h_n for which it is possible to give the asymptotically optimal value

$$h_{opt} = \left(\frac{2\pi K_0 C_\theta}{L_2 \lambda^2 \text{vol}(B) \text{vol}(B')} \right)^{1/4}, \quad (2.4)$$

where the constants $C_\theta, K_0, L_2 > 0$ depend on f (c.f. [1, Remark 4.4] and [10, Remark 4.3.3]).

Due to the complex dependency structure of the random field $\{\log_{B'+Y_i}(\xi_i)\}_{i \geq 1}$, it is necessary to modify the estimator in order to study its asymptotic normality. Namely, we consider

$$\hat{\mathcal{E}}_f^*(B) := -\frac{1}{\lambda \text{vol}(B)} \sum_{i \geq 1} \mathbb{1}_{\{Y_i^* \in B\}} \log \hat{f}_{B'+Y_i^*}(\xi_i^*), \quad (2.5)$$

where $\Psi_\lambda^* = \{(Y_i^*, \xi_i^*)\}_{i \geq 1}$ is an independent copy of the original MPPP Ψ_λ . It is shown in [1, Theorem 5.7] that under suitable assumptions, in particular when the sequences of growing windows $\{B_n\}_{n \geq 1}$ and $\{B'_n\}_{n \geq 1}$ satisfy $B_n = [0, b_n]^3$, $B'_n = [0, m_n]^3$ and $b_n = m_n^{4+\delta}$ for some $\delta > 0$, the central limit theorem

$$\sqrt{\text{vol}(B_n)} \frac{\hat{\mathcal{E}}_f^*(B_n) - \hat{\mu}_{B_n}}{\sigma_n} \xrightarrow{d} \mathcal{N}(0, 1) \quad (2.6)$$

holds. Here,

$$\begin{aligned} \hat{\mu}_{B_n} &= \frac{\#(\Pi_\lambda \cap B_n)}{\lambda \text{vol}(B_n)} \mathbb{E}[-\log \hat{f}_{B'_n}(\xi_0^*)], \\ \sigma_n^2 &= \text{Var}(\log \hat{f}_{B'_n}(\xi_0^*)) + \lambda^2 \int_{B'_n} \text{Cov}(\log \hat{f}_{B'_n}(\xi_0^*), \log \hat{f}_{B'_n}(\xi'_y)) dy, \end{aligned} \quad (2.7)$$

where $\{\xi'_y, y \in \mathbb{R}^3\}$ are i.i.d. copies of ξ_0 and $\#B$ is the cardinality of a finite set B . A simulation study of this limit theorem is presented in Section 4.

We would like to remark that the results mentioned in this section also hold for MPPPs whose marks ξ_i take values in more abstract spaces. Thus, the marks need not represent the direction but any other characteristic of a fibre, as long as its distribution has a density satisfying certain constraints. These are nevertheless fulfilled by distributions of many quantities of interest, such as fibre length or curvature.

3 Detection of inhomogeneities. Methodology.

How can a fibre cluster in a FRP be detected? These kind of inhomogeneities arise for instance when the production process fails to orient the fibres of a FRP properly. This section proposes an entropy-based method to answer this question that relies on the entropy of the directional distribution of the fibres in view of the ability of entropy to

perceive abrupt changes in a distribution. The aforementioned estimator $\widehat{\mathcal{E}}_f^*(B)$ and its asymptotic properties will thus play a decisive role in this procedure.

An important necessary condition for a successful performance will be that the directional distribution of the fibres in and outside the inhomogeneity region strongly differ from each other. Further possible limitations of our method will be discussed at the end of the section.

3.1 Description of the method

Let $W \subset \mathbb{R}_+^3$ denote the window (piece of material) to be analysed, $A \subset W$ the inhomogeneity region, and $B \subset \mathbb{R}_+^3$ the scanning window. For simplicity, these sets will be assumed to be cubes of side lengths $w, a, b > 0$ respectively, i.e. $W = [0, w]^3$, $A = [0, a]^3 + a_0$ for some $a_0 \in \mathbb{R}_+^3$, and $B = [0, b]^3$. Moreover, we will assume that $0 < b < a < w$. In particular, $b < a$ means that the volume of the scanning window is smaller than the volume of the inhomogeneity region. Notice that a scanning window that is significantly bigger than the inhomogeneity will fail to identify that region. Moreover, the simulation study has indicated this assumption to provide the best performance (see e.g. Figure 8).

The presence of A will be detected by means of hypothesis testing. Under the hypothesis of homogeneity, the directional distribution is the same for any fibre, and we will assume that this distribution has a density f . Rejecting the hypothesis will be an indicator of the existence of A .

In a first step, the window W will be scanned using the window B by letting it run through the points in the minus-sampled window $W \ominus B$ to avoid boundary effects. Recall that given two sets $B_1, B_2 \subset \mathbb{R}^3$, their *Minkowski difference* $B_1 \ominus B_2$ is defined as the set $\{x - y : x \in B_1, y \in B_2\}$.

At each point $x \in W \ominus B$, the entropy will be locally estimated in $B + x$, generating the stationary random field

$$\{\widehat{\mathcal{E}}_f^*(B + x), x \in W \ominus B\}.$$

For a finite number of observation points $x_1, \dots, x_n \in W \ominus B$, we set

$$\begin{aligned} \tilde{\mu}(\{x_1, \dots, x_n\}) &:= \text{median}\{\widehat{\mathcal{E}}_f^*(B + x_1), \dots, \widehat{\mathcal{E}}_f^*(B + x_n)\}, \\ \hat{\mu}(\{x_1, \dots, x_n\}) &:= \frac{1}{n} \sum_{i=1}^n \widehat{\mathcal{E}}_f^*(B + x_i), \\ \hat{\sigma}^2(\{x_1, \dots, x_n\}) &:= \frac{1}{n-1} \sum_{i=1}^n (\widehat{\mathcal{E}}_f^*(B + x_i) - \hat{\mu}(\{x_1, \dots, x_n\}))^2. \end{aligned}$$

In order to quantify a substantial deviation from the hypothesis of homogeneity, we use in a second step the asymptotic Gaussianity of $\widehat{\mathcal{E}}_f^*(B)$ stated in (2.6) to view $\{\widehat{\mathcal{E}}_f^*(B +$

$x), x \in W \ominus B\}$ approximately as a Gaussian random field (GRF). To each of its elements, the so-called 3σ -rule may be applied, which states that for a Gaussian random variable X it holds that

$$\mathbb{P}(|X - \mathbb{E}[X]| > 3\sqrt{\text{Var}(X)}) \approx 0.0027. \quad (3.1)$$

This result can be extended to arbitrary distributions with unimodal density, its mode as center, and the upper bound 0.05 instead of 0.0027 in (3.1), see [6] for a review. This rule underlies our proposed method to detect the inhomogeneity region A .

Recall that the direction of the fibres lying inside the region A is supposed to strongly differ from the direction of the fibres outside of A . If the value of the local entropy $\widehat{\mathcal{E}}_f^*(B+x)$ substantially deviates from the “median entropy” $\tilde{\mu}(\{x_1, \dots, x_n\})$, this will indicate that the hypothesis is violated, suggesting that the point x lies in A . The employment of the median rather than the empirical mean is our proposal to avoid outliers. The estimated inhomogeneity region will thus be defined as the excursion set

$$\widehat{A}_{B,W}^{(n)} := \{x \in W \ominus B : |\widehat{\mathcal{E}}_f^*(B+x) - \tilde{\mu}_n| > 3\hat{\sigma}_n\},$$

where $\tilde{\mu}_n = \tilde{\mu}(\{x_1, \dots, x_n\})$ and $\hat{\sigma}_n = \hat{\sigma}^2(\{x_1, \dots, x_n\})$ with $x_1, \dots, x_n \in W \ominus B$. In practice, we will take $\tilde{\mu} = \tilde{\mu}((W \ominus B) \cap r\mathbb{Z}^3)$ and $\hat{\sigma} = \hat{\sigma}((W \ominus B) \cap r\mathbb{Z}^3)$, where $r\mathbb{Z}^3$ denotes the lattice of mesh size $r > 0$. In simulations, these lattice points will also be used to discretize $W \ominus B$.

3.2 Scanning window size

A significant issue concerning the proposed method is the choice of the size of the scanning window B . On the one hand, B should be large enough so that the local entropy is estimated accurately. On the other hand, a far too large scanning window may oversee the inhomogeneity.

In this paragraph we give an answer to this question under the condition that some a priori information about the side length relation between the inhomogeneity and the whole observation window is known. In particular we will see that, if these lengths depend linearly on each other, then the side length of the observation window B is a multiple of the side length of the inhomogeneity region A and the factor will depend on the quantile $\mathbb{P}(|\widehat{\mathcal{E}}_f^*(B) - \tilde{\mu}| > 3\hat{\sigma})$. Setting this quantile to 0.05 and if for instance the side length w of the observation window and the side length a of the inhomogeneity satisfy the relation $w = 7a$, then B will be chosen to have side length $b = 0.489a$.

We use the notion of *expected distance in measure* to quantify the approximation error of the estimated (random) inhomogeneity region $\widehat{A}_{B,W}$. For an abstract measurable space (M, ν) and any two measurable sets $S_1, S_2 \subseteq M$, the *distance in measure* between them is given by the measure of their symmetric difference $S_1 \triangle S_2 := (S_1 \cap S_2^c) \cup (S_1^c \cap S_2)$. The analogous concept for random sets is known as the *expected distance in measure*, which is defined as follows.

Definition 3.1. Let (M, ν) be a measurable space. The expected distance in measure with respect to ν is the function $d_\nu: \mathcal{B}(M) \times \mathcal{B}(M) \rightarrow [0, \infty]$ defined as

$$d_\nu(\Gamma_1, \Gamma_2) := \mathbb{E}[\nu(\Gamma_1 \triangle \Gamma_2)]$$

for any Borel measurable random closed sets $\Gamma_1, \Gamma_2: \Omega \rightarrow \mathcal{B}(M)$.

In view of this definition, the size, i.e. the Lebesgue measure or volume, of the scanning window B is optimal when the expected distance in measure between the original inhomogeneity region A and the estimated region $\hat{A}_{B,W}$ is minimal. In other words, B will be chosen to satisfy

$$B = \operatorname{argmin}_{\tilde{B}} d_{\text{vol}}(A, \hat{A}_{\tilde{B},W}).$$

Notice that the MPPP $\Psi_{\lambda,A} := \{(Y_i, \xi_i)\}_{i \geq 1}$ modeling a fibre system with an inhomogeneity has marks ξ_i that are independent but *not* identically distributed. Namely, marks that correspond to points in A follow a directional distribution with density g whereas marks corresponding to points in $W \setminus A$ have a different direction with distribution density f . In this case, the entropy estimator will be simply denoted by $\hat{\mathcal{E}}^*$ since there is no specific density distribution to refer to.

For simplicity of the calculations and to avoid further boundary effects, we will assume that the observation window W is large enough and the inhomogeneity region A small enough so that

$$A \oplus B \subseteq W \ominus B. \quad (3.2)$$

The expected distance in measure with respect to the measure $\text{vol}(\cdot)$ has the following expression.

Lemma 3.2. Let $R_1 = (A \oplus B) \setminus A$ and let $R_2 = A \setminus (A \ominus B)$, where ∂A denotes the boundary of A . Then,

$$\begin{aligned} d_{\text{vol}}(A, \hat{A}_{B,W}) &= \text{vol}(A) + \text{vol}((W \ominus B) \setminus (A \oplus B)) \mathbb{P}(|\hat{\mathcal{E}}_f^*(B) - \tilde{\mu}| > 3\hat{\sigma}) \\ &\quad - \text{vol}(A \ominus B) \mathbb{P}(|\hat{\mathcal{E}}_g^*(B) - \tilde{\mu}| > 3\hat{\sigma}) \\ &\quad + \int_{R_1} \mathbb{P}(|\hat{\mathcal{E}}^*(B+x) - \tilde{\mu}| > 3\hat{\sigma}) dx - \int_{R_2} \mathbb{P}(|\hat{\mathcal{E}}^*(B+x) - \tilde{\mu}| > 3\hat{\sigma}) dx. \end{aligned}$$

Proof. On the one hand, by definition of expected distance and symmetric difference we have that

$$d_{\text{vol}}(A, \hat{A}_{B,W}) = \text{vol}(A) + \mathbb{E}[\text{vol}(\hat{A}_{B,W})] - 2\mathbb{E}[\text{vol}(A \cap \hat{A}_{B,W})]. \quad (3.3)$$

On the other hand, Fubini's theorem yields

$$\mathbb{E}[\text{vol}(\hat{A}_{B,W})] = \int_{W \ominus B} \mathbb{P}(|\hat{\mathcal{E}}^*(B+x) - \tilde{\mu}| > 3\hat{\sigma}) dx$$

and since $A \subseteq W \ominus B$ by (3.2), this also yields

$$\mathbb{E}[\text{vol}(A \cap \hat{A}_{B,W})] = \int_A \mathbb{P}(|\hat{\mathcal{E}}^*(B+x) - \tilde{\mu}| > 3\hat{\sigma}) dx.$$

Let us analyze the first integral. Notice that if $x \in (W \ominus B) \setminus (A \oplus B)$, then $B+x$ does not intersect A and therefore $\hat{\mathcal{E}}^*(B+x) = \hat{\mathcal{E}}_f^*(B+x)$. On the other hand, if $x \in A \ominus B$, then $\hat{\mathcal{E}}^*(B+x) = \hat{\mathcal{E}}_g^*(B+x)$. In view of the stationarity of $\{\hat{\mathcal{E}}_f^*(B+x), x \in \mathbb{R}^3\}$ and $\{\hat{\mathcal{E}}_g^*(B+x), x \in \mathbb{R}^3\}$ we have that

$$\begin{aligned} & \int_{W \ominus B} \mathbb{P}(|\hat{\mathcal{E}}^*(B+x) - \tilde{\mu}| > 3\hat{\sigma}) dx \\ &= \text{vol}((W \ominus B) \setminus (A \oplus B)) \mathbb{P}(|\hat{\mathcal{E}}_f^*(B) - \tilde{\mu}| > 3\hat{\sigma}) \\ &+ \text{vol}(A \ominus B) \mathbb{P}(|\hat{\mathcal{E}}_g^*(B) - \tilde{\mu}| > 3\hat{\sigma}) + \int_{R_1 \cup R_2} \mathbb{P}(|\hat{\mathcal{E}}^*(B+x) - \tilde{\mu}| > 3\hat{\sigma}) dx. \end{aligned}$$

Analogously we obtain

$$\begin{aligned} & \int_A \mathbb{P}(|\hat{\mathcal{E}}^*(B+x) - \tilde{\mu}| > 3\hat{\sigma}) dx \\ &= \text{vol}(A \ominus B) \mathbb{P}(|\hat{\mathcal{E}}_g^*(B) - \tilde{\mu}| > 3\hat{\sigma}) + \int_{R_2} \mathbb{P}(|\hat{\mathcal{E}}^*(B) - \tilde{\mu}| > 3\hat{\sigma}) dx. \end{aligned}$$

Plugging these equalities into (3.3) leads to the desired expression. \square

This result can be applied in order to obtain an upper bound for the distance $d_{\text{vol}}(A, \hat{A}_{B,W})$ in terms of the side lengths of W, A and B . The side length of the scanning window B will be chosen to minimize this bound and it will depend on the side lengths of W and A .

Taking into account assumption (3.2), the inhomogeneity region can be expressed as $A = [0, a]^3 + a_0$ for some $a > 0$ and $a_0 \in W \setminus B''$ with $B'' = [0, 2b+a]^3$. In this case, the following identities hold

$$W \ominus B = [0, w-b]^3, \quad A \ominus B = [0, a-b]^3 + a_0, \quad A \oplus B = [0, a+b]^3 + a_0.$$

Further, set

$$\alpha_f := \mathbb{P}(|\hat{\mathcal{E}}_f^*(B) - \tilde{\mu}| > 3\hat{\sigma}), \quad \text{and} \quad \alpha_g := \mathbb{P}(|\hat{\mathcal{E}}_g^*(B) - \tilde{\mu}| > 3\hat{\sigma}).$$

In view of Lemma 3.2 and since $\text{vol}(R_1) = \text{vol}(A \oplus B) - \text{vol}(A)$,

$$\begin{aligned} d_{\text{vol}}(A, \hat{A}_{B,W}) &\leq \text{vol}(A) + \text{vol}((W \ominus B) \setminus (A \oplus B))\alpha_f - \text{vol}(A \ominus B)\alpha_g + \text{vol}(R_1) \\ &= \text{vol}(A \oplus B)(1 - \alpha_f) + \text{vol}(W \ominus B)\alpha_f - \text{vol}(A \ominus B)\alpha_g \\ &= (a+b)^3(1 - \alpha_f) + (w-b)^3\alpha_f - (a-b)^3\alpha_g. \end{aligned}$$

Furthermore, the entropy of the directional distribution of fibres lying in A must deviate from the median $\tilde{\mu}$ with higher probability than when the fibres lie outside of A . Hence, we can assume that $\alpha_f < \alpha_g$ and therefore get

$$d_{\text{vol}}(A, \hat{A}_{B,W}) \leq (a+b)^3(1-\alpha_f) + ((w-b)^3 - (a-b)^3)\alpha_f.$$

With the help of computational software (here Mathematica was used) the value of

$$\operatorname{argmin}_b \{(a+b)^3(1-\alpha_f) + ((w-b)^3 - (a-b)^3)\alpha_f\}$$

is given by the expression

$$b_{\text{opt}} = \frac{\sqrt{\alpha_f(w-a)(w+(3-4\alpha_f)a)} - (1-2\alpha_f)a - \alpha_f w}{1-\alpha_f}. \quad (3.4)$$

Since $a < w$ and α_f small, it holds that $\alpha_f(w-a)(w+(3-4\alpha_f)a) > 0$. Yet, the parameter α_f and the relation between a and w must be chosen in order to assure $0 < b_{\text{opt}} < a$. As mentioned at the beginning of this paragraph and following the 3σ -rule, setting $\alpha_f = 0.05$ we chose $w = 7a$ for the simulation study, for which (3.4) yields $b = 0.489a$.

4 Simulations

In order to validate the estimators introduced in Section 2 as well as to examine the efficiency of the inhomogeneity detection method presented Section 3, this section is devoted to several tests on simulated data. One of the main reasons for working with simulations is that the input is known and therefore the estimation error can be computed accurately. Moreover, the MPPPs considered for these simulations model structures with a high fibre intensity, for which the analysis of real μCT data is usually difficult and costly. The high intensity plays an important role in the estimation of the entropy and therefore in the detection method: since the estimators \hat{f}_B and $\hat{\mathcal{E}}^*$ are asymptotically consistent, the amount of data (fibres) has to be sufficiently large in order to provide a good approximation.

4.1 Density estimation

In order to test the density estimator \hat{f}_B defined in (2.2), an independently marked MPPP of intensity 15 was simulated in an observation window $B = [0, 50]^3$. Table 1 and Figure 2 display the approximation error of \hat{f}_B for different kernels (see e.g. [13] for definitions and further details about kernel functions). Here, the computed error is given by $\max_{x \in S_d^2} |\hat{f}_B(x) - f(x)|$, where S_d^2 denotes a discretization of the sphere into a grid of points. Since the constants appearing in the expression (2.4) of the optimal bandwidth cannot be computed explicitly, by empirical observation the bandwidth was

set to be $h = \left(\frac{1+\text{vol}(B)}{\text{vol}(B)^{10/9}}\right)^{1/4}$. Notice that this quantity has the same asymptotic order as (2.4) when $\text{vol}(B') = \text{vol}(B)^{1/9}$.

	Uniform	Schlادitz(2)	Watson(2)	Fisher(2)
Biweight	0.0094	0.0125	0.2487	0.0152
Epanechnikov	0.0195	0.0218	0.3393	0.0499
Triangular	0.0089	0.0089	0.2751	0.0226
Tricube	0.0082	0.0098	0.2536	0.0149
Triweight	0.0190	0.0285	0.2487	0.0521
Uniform	0.0524	0.0654	0.5284	0.1565

Table 1: Approximation error of different distribution densities f by its kernel density estimator \hat{f}_B for different kernels measured in the uniform convergence metric.

The density f corresponds in each case to the uniform, Schlادitz(2), Watson(2) or Fisher(3) directional distribution on the sphere, see e.g. [2, 9] for precise definitions of these distributions.

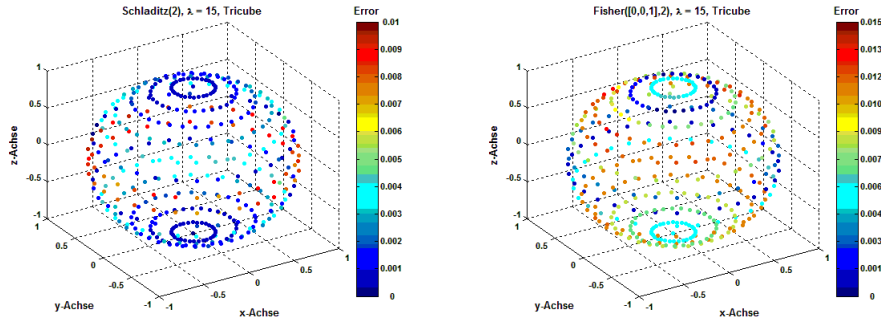


Figure 2: Graphical visualization of the estimation error for the directional distributions Schlادitz(2) and Fisher([0, 0, 1], 2) with Tricube kernel.

An exhaustive simulation study with more distributions, kernels and observation window sizes can be found in [10].

4.2 Entropy estimator

For the study of the entropy estimator defined in (2.3), an independently marked MPPP of intensity 15 was simulated in an observation window $B = [0, 50]^3$. Recall that the density f is estimated in the additional window $B' = [0, 4]^3$. Although B' could still have been chosen to equal the whole observation window B at this point, we took on account as far as possible the side-length relation between these windows appearing in the asymptotic normality test based on (2.6). The window B' being small, it has

to contain sufficiently many points (fibres) to obtain a fairly good estimation. This is guaranteed by the fact that the intensity of the underlying point process is relatively high.

In the subsequent testing, the kernel K is fixed to tricube because the previous simulations indicated it to yield the best approximation. The bandwidth is the same as for the density estimation, $h = \left(\frac{1+\text{vol}(B)}{\text{vol}(B)^{10/9}}\right)^{1/4}$, with the same asymptotic order as (2.4) and $\text{vol}(B') = \text{vol}(B)^{1/9}$.

	\mathcal{E}_f	$\overline{\hat{\mathcal{E}}_f(B)}$	$\text{Var } \hat{\mathcal{E}}_f(B)$	$\ Err\ _\infty$	MSQE
Uniform	2.5310	2.5165	8.7105e-07	0.0157	0.0459
Schladitz(2)	2.3554	2.3525	9.0843e-07	0.0039	0.0067
Fisher(2)	1.7239	1.8930	2.3662e-06	0.1715	0.3782
Watson(2)	1.8646	1.6849	2.8397e-07	0.1804	0.5682

Table 2: True value, arithmetic mean, sample variance, absolute error and mean square error of the entropy estimation.

The asymptotic normality of the modified estimator $\hat{\mathcal{E}}_f^*(B)$ introduced in (2.5) has been tested by running 400 simulations of $\hat{\mathcal{E}}_f^*(B)$ based on two independent copies, Ψ_{20} and Ψ_{20}^* , of an independently marked MPPPs of intensity 20 observed in a window $B = [0, 30]^3$. The auxiliary window where the density is estimated is $B' = [0, 4]^3$. Figure 3 summarizes the results for different directional distributions.

In the computation of the normalization terms in (2.7), the mean has been replaced by its empirical estimation obtained from 180 independent realizations of $-\log \hat{f}_{B'}(\xi_0^*)$. The integral in the covariance term has been discretized using 343 lattice points in B' and the corresponding empirical estimator of each covariance term.

The size of the observation windows and the high density of the underlying point process reproduce the asymptotic nature of the theoretic result.

4.3 Inhomogeneity detection

In this section we present two examples where our proposed method detects a (self-generated) inhomogeneity region. The size relationship between the observation window, the inhomogeneity region and the scanning window have been determined following the optimal expression in (3.4) with $\alpha_f = 0.05$ and $w = 7a$.

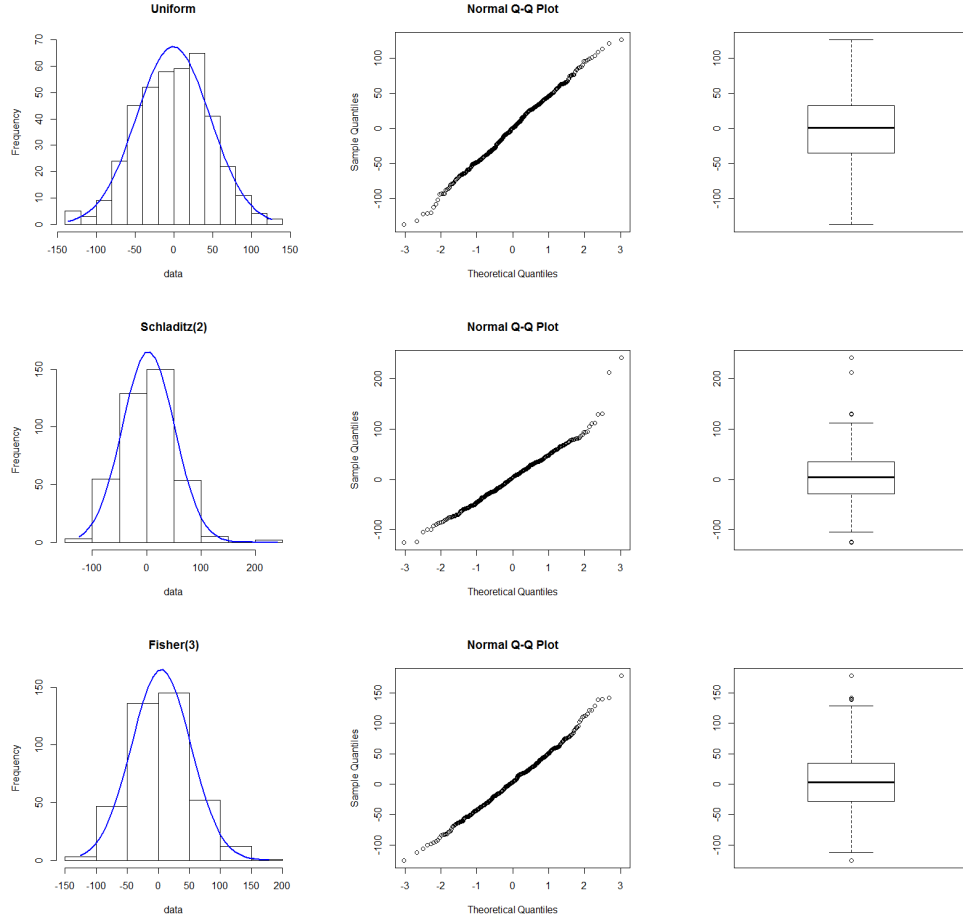


Figure 3: Fitted histogram, q-q plot and box plot for the entropy estimator $\mathcal{E}_f^*(B)$ with uniform distribution, Schladitz(2)- and Fisher(3)-distribution, respectively.

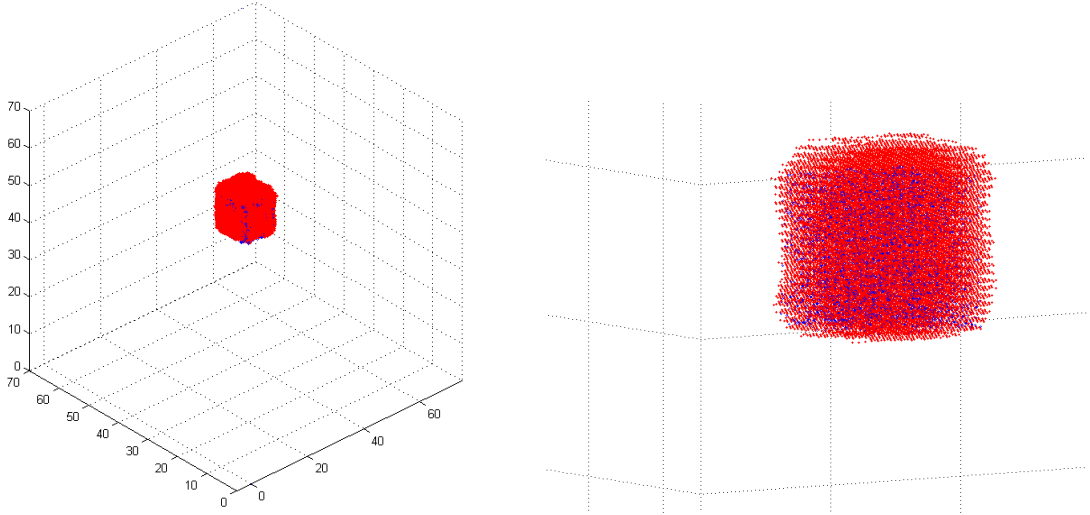


Figure 4: The cube where the inhomogeneity resides is completely recognized.

In the first example, an underlying independently marked MPPP Ψ_5 of intensity 5 is generated in the window $W = [0, 70]^3$. The inhomogeneity region, colored in blue in Figure 4, is the cube $A = [0, 10]^3 + (35, 35, 35)$. W is scanned by running the window $B = [0, 5]^3$ over 130^3 lattice points. The fibres (points) inside A have directions following a uniform distribution, whereas the direction of fibres outside the inhomogeneity region is Fisher(10)-distributed. The red points constitute the (discretized) estimated inhomogeneity region $\hat{A}_{B,W}$ obtained as described in Section 3.

The second example illustrates that this method also works when there are more than just one inhomogeneity region, as long as each component satisfies the constraints (3.2) and (3.4). In this case, $A = A_1 \cup A_2$ with $A_1 = [0, 10] + (10, 10, 10)$ and $A_2 = [0, 10] + (30, 30, 30)$. The direction of the fibres inside the inhomogeneity region A follow a uniform distribution, whereas the direction of fibres outside A have Fisher(10)-distributed directions.

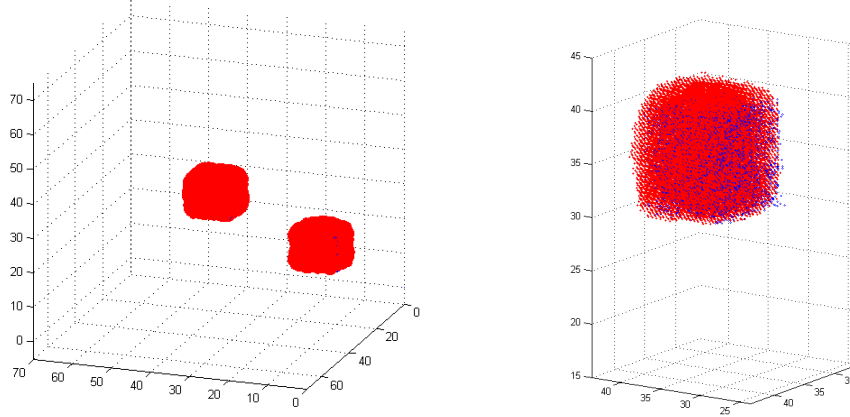


Figure 5: Two cubes with marks differently distributed are also completely recognized.

5 Discussion and outlook

Inhomogeneities such as fibre clusters or deformations in FRPs are directly related to changes in the direction of the fibres that compose the material and detecting these regions is of special interest in order to control its production process.

In this paper we propose a method to detect inhomogeneities based on the entropy as a measure of change in the directional distribution of the fibres. One major advantage of using the entropy is that the problem reduces to a univariate change-point detection that can be performed via the 3σ -rule due to the asymptotic normality of the entropy estimator.

The tests presented in Section 4 support the applicability of the theoretical estimator for the density and entropy of directional distributions that underlie the method exposed in Section 3. Within the constraints concerning the shape of the observation windows W , B and the inhomogeneity region A , as well as assuming prior knowledge about the size relation between W and A , it is possible to derive an optimal size for the scanning window B .

A finer analysis of the output reveals some discrepancies along the boundary of the inhomogeneity A as Figure 6 shows. The reason for this is that, whenever the lower-left corner of the scanning window B is near the boundary ∂A but still outside of A , and $B \cap A$ covers most of B , the entropy on B is almost the same as when B completely lies in A . Thus, even if the whole scanning window does not lie inside A , the estimated entropy $\hat{\mathcal{E}}^*(B)$ will be considered an outlier.

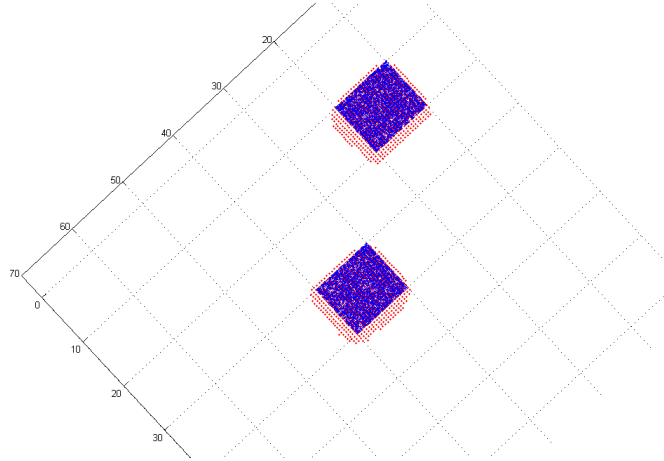


Figure 6: Boundary effect in the inhomogeneity detection.

The shape of the inhomogeneity A seems to affect the performance as well. On the one hand, it is not anymore possible to establish a simple relation between A and the scanning window B (as long as it has a shape of different nature). On the other hand, stronger boundary effects should be expected. In Figure 7 the inhomogeneity has been chosen to be a ball of radius 5 centered in the observation window, and the scanning window is been kept a cube of side-length 5. We observe how in this case the inhomogeneity is still well recognized, however additional points appear, and the boundary effects observed in the previous example increase.

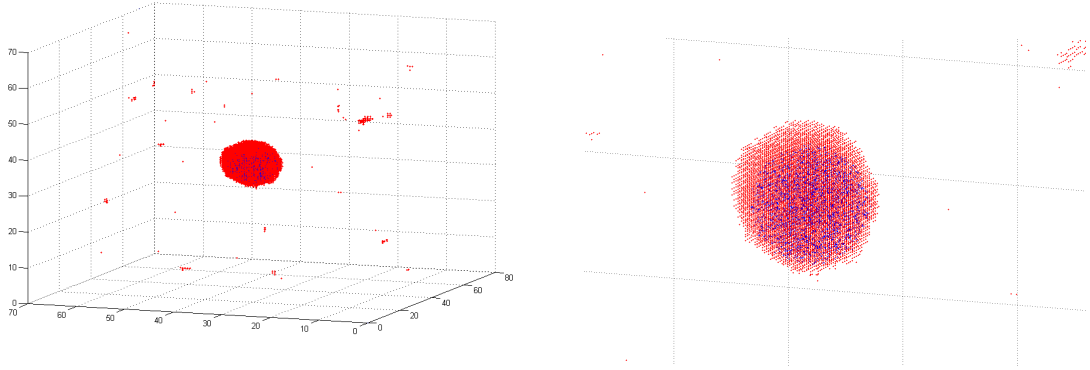


Figure 7: The method becomes less efficient if the form of the inhomogeneity and of the scanning window notably differ. Boundary effects appear as well.

Further issues appear when the size of the scanning window B is much smaller or larger than the size of the inhomogeneity A , c.f. Figure 8.

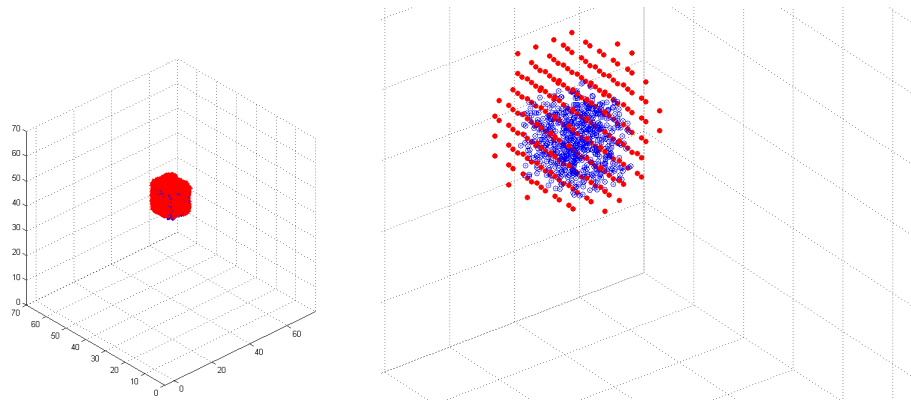


Figure 8: The performance also decays when the inhomogeneity region is actually larger than expected (left) or smaller (right).

In conclusion, these investigations reveal our method to be a good detection procedure that may potentially be used for a first fast rough scan to rule out substantial inhomogeneities, where their contours need not be perfectly recognized.

References

- [1] P. Alonso-Ruiz and E. Spodarev, *Estimation of entropy for Poisson marked point processes*, Advances in Applied Probability **49** (2017), no. 1, to appear.
- [2] N. I. Fisher, T. Lewis, and B. J. J. Embleton, *Statistical Analysis of Spherical Data*, Cambridge University Press, Cambridge, 1993, Revised reprint of the 1987 original.
- [3] J. Li, X. Zhang, and D. R. Jeske, *Nonparametric multivariate CUSUM control charts for location and scale changes*, J. Nonparametr. Stat. **25** (2013), no. 1, 1–20.
- [4] I. B. MacNeill and V. K. Jandhyala, *Change-point methods for spatial data*, Multivariate environmental statistics, North-Holland Ser. Statist. Probab., vol. 6, North-Holland, Amsterdam, 1993, pp. 288–306.
- [5] J. Ohser and K. Schladitz, *3D Images of Materials Structures: Processing and Analysis*, Wiley, Weinheim, 2009.
- [6] F. Pukelsheim, *The three sigma rule*, Amer. Statist. **48** (1994), no. 2, 88–91.
- [7] C. Redenbach, A. Rack, K. Schladitz, O. Wirjadi, and M. Godehardt, *Beyond imaging: on the quantitative analysis of tomographic volume data*, Int. J. of Materials Research **103** (2012), no. 2, 217 – 227.
- [8] C. Redenbach, K. Schladitz, I. Vecchio, and O. Wirjadi, *Image analysis for microstructures based on stochastic models*, GAMM-Mitteilungen **37** (2014), no. 2, 281–305.

- [9] K. Schladitz, S. Peters, D. Reinel-Bitzer, A. Wiegmann, and J. Ohser, *Design of acoustic trim based on geometric modeling and flow simulation for non-woven*, Computational Materials Science **38** (2006), no. 1, 56 – 66.
- [10] J. Schwarz, *Schätzer für Dichte und Entropie der Markenverteilung eines markierten Poisson-Punktprozesses auf der Sphäre*, Bachelor thesis, Ulm University, 2016.
- [11] C. E. Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423, 623–656.
- [12] L. Shu, W. Jiang, and K. Tsui, *A standardized scan statistic for detecting spatial clusters with estimated parameters*, Naval Res. Logist. **59** (2012), no. 6, 397–410.
- [13] A. B. Tsybakov, *Introduction to Nonparametric Estimation*, Springer Series in Statistics, Springer, New York, 2009.